

A method for capitalizing text in a document includes processing a reference corpus to construct a plurality of dictionaries of capitalized terms, where the plurality of dictionaries include a singleton dictionary and a phrase dictionary. Each record in the singleton dictionary contains a word in lowercase, a range of phrase lengths $m:n$ for capitalized phrases that the word begins, where m is a minimum phrase length and n is a maximum phrase length, and where each record in the phrase dictionary includes a multi-word phrase in lowercase. The method adds proper capitalization to an input monospace document by capitalizing words found in mandatory capitalization positions; and by looking up each word in the singleton dictionary and, if the word is found in the singleton dictionary, testing the corresponding phrase length range. If the phrase length range indicates that the word does not start a multi-word phrase, the method capitalizes the word, while if the phrase length range indicates that the word does start a multi-word phrase, the method tests the word and an indicated plurality of next words as a candidate phrase to determine if the candidate phrase is found in the phrase dictionary and, if it is, capitalizes the words of the multi-word phrase. If the candidate phrase is not found in the phrase dictionary, the method changes the number of words in the candidate phrase (e.g., decrements by one) to form a revised candidate phrase, and determines whether the revised candidate phrase is found in the phrase dictionary.